

# INFORMATION CONTENT IN THE GALAXY ANGULAR POWER SPECTRUM FROM THE SLOAN DIGITAL SKY SURVEY AND ITS IMPLICATION ON WEAK LENSING ANALYSIS

JOUNGHUN LEE

Department of Physics and Astronomy, FPRD, Seoul National University, Seoul 151-747, Korea : jounghun@astro.snu.ac.kr

UE-LI PEN

Canadian Institute for Theoretical Astrophysics, Toronto, ON M5S 3H8, Canada: pen@cita.utoronto.ca  
 Draft version August 27, 2008

## ABSTRACT

We analyze the photometric redshift catalog of the Sloan Digital Sky Survey Data Release 5 (SDSS DR5) to estimate the Fisher information in the galaxy angular power spectrum with the help of the Rimes-Hamilton technique. It is found that the amount of Fisher information contained in the galaxy angular power spectrum is saturated at lensing multipole scale  $300 \leq l \leq 2000$  in the redshift range  $0.1 \leq \text{photo-}z < 0.5$ . At  $l = 2000$ , the observed information is two orders of magnitude lower than the case of Gaussian fluctuations. This supports observationally that the translinear regime of the density power spectrum contains little independent information about the initial cosmological conditions, which is consistent with the numerical trend shown by Rimes-Hamilton. Our results also suggest that the Gaussian-noise description may not be valid in weak lensing measurements.

*Subject headings:* cosmology:theory — methods:statistical — large-scale structure of universe

## 1. INTRODUCTION

Cosmic shear, which refers to the weak gravitational lensing by large-scale structure in the universe, is an invaluable tool in cosmology since it directly probes the gravitational clustering of dark matter in the universe without concerns about light-to-matter bias. Very recently, the three dimensional map of dark matter distribution has been indeed reconstructed on cosmological scales from the measurements of cosmic shear (Massey et al. 2007), which marks an advent of *lensing cosmology*. Among many lensing observables, the angular power spectrum is regarded particularly important, since it is in principle capable of constraining the key cosmological parameters with high precision, including the enigmatic nature of dark energy (see Hoekstra & Jain 2008, for a recent review).

The success of the angular power spectrum as a cosmological probe, however, is subject to one critical issue: how much information does it preserve about the initial conditions of the universe? A natural expectation is that some but not all information might have been destroyed in the subsequent nonlinear evolution. The amount of information contained in the power spectrum about the initial cosmological conditions is directly related to the precision of the cosmological parameters constrained by using the matter power spectrum. In other words, the non-Gaussian errors caused by the loss of information in the power spectrum would propagate into large uncertainties in the determination of the cosmological parameters. The error propagation and precision on cosmological parameters can be quantified in terms of the Fisher information contained in the matter power spectrum (Tegmark et al. 1997).

Rimes & Hamilton (2005, hereafter, RH05) have for the first time estimated the Fisher information in the matter power spectrum. Basically, RH05 have calculated the information content about the amplitude of the lin-

ear power spectrum averaged over an ensemble of many N-body realizations and found very little independent information at the translinear scale. Somewhat surprisingly, however, RH05 also found that there is a sharp rise in the amount of information in the nonlinear scale (see also Rimes & Hamilton 2006; Rimes et al. 2006). This phenomenon has been also noted by the analytic work of Neyrinck et al. (2006).

Yet, according to the recent results derived by Neyrinck & Szapudi (2007) based on the halo model, the Fisher information in the density power spectrum about all key cosmological parameters including the initial amplitude is highly degenerate both in the translinear and the nonlinear regime. Their work has indicated that it might not be possible to extract the initial cosmological conditions from the nonlinear dark matter power spectrum to a high statistical accuracy.

Thus, the previous numerical and analytic results forecast that the accuracy in the determination of the cosmological parameters in lensing cosmology may be lower due to the non-Gaussian errors. As the next generation of large surveys optimized for weak lensing will soon be on the pipeline, it is imperative to test observationally the information content in the lensing power spectrum. In this Letter we attempt to do this test by applying the RH05 technique to the photometric galaxy catalogs from the Sloan Digital Sky Survey Data Release 5 (SDSS DR5, Adelman-McCarthy et al. 2007) at typical lens redshifts  $z \sim 0.3$ .

## 2. DATA

We use the Photoz2 and the PhotoPrimary catalog which are both publicly available at the web site of the SDSS DR5 (<http://www.sdss.org/dr5>). The Photoz2-catalog contains information on galaxy photometric redshift (photo- $z$ ,  $z_p$ ), while the PhotoPrimary catalog has imaging data of galaxy right ascension ( $\alpha$ ), declination ( $\delta_I$ ), and dereddened model magnitude ( $M_r$ ). For our

TABLE 1

photo- $z$	$\delta_I$	$\alpha$	# of galaxies	# of cells
[0.1, 0.2)	[0°, 60°)	[120°, 240°)	2794590	55
[0.2, 0.3)	[0°, 60°)	[120°, 240°)	4865953	55
[0.3, 0.4)	[0°, 60°)	[120°, 240°)	9768987	55
[0.4, 0.5)	[0°, 60°)	[120°, 240°)	9500664	55

analysis, we select those galaxies in the catalog whose dereddened model magnitude, photo- $z$  and angular positions satisfy the constraints of  $M_r < 22.0$ ;  $0.0 \leq z_p < 1$ ;  $0^\circ \leq \delta_I < 60^\circ$ ;  $120^\circ \leq \alpha < 240^\circ$ .

Selected are a total of 42587179 galaxies which we divide into smaller samples according to their angular positions and photo- $z$ 's. We first decompose the selected portion of the sky into 55 cells, each having approximately same linear size of  $10^\circ$ . To make equal-size cells on *non-flat* sky, the bin size of  $\delta_I$  is set at the constant value of  $\delta_I = 10^\circ$ , while the bin size of  $\alpha$  is adjusted by a factor of  $1/\cos\delta_{Im}$  to compensate for the decrease in the width of  $\Delta\alpha$  with  $\delta_I$ . Here, we choose  $\delta_{Im}$  as the median value of  $\delta_I$  at each cell. We end up with having a total of 55 cells on the plane of the sky in the selected angular range. Table 1 lists the ranges of  $\alpha$  and  $\delta_I$ , a total number of galaxies ( $N_g$ ) and cells ( $N_C$ ) in four photo- $z$  slices.

We treat approximately each cell as a square region and decompose it further into  $128^2$  pixels of equal size. Counting the number of galaxies at each pixel, we find those pixels where no galaxy is located and mask them and their nearest neighbor pixels as gap regions, assigning zero density contrast. After masking the gap regions, we bin the selected photo- $z$  range with bin size of  $\Delta z_p = 0.1$ . A set of 55 cells in each redshift slice represents an ensemble of realizations for the calculation of the galaxy angular power spectrum. The size of the field is well in linear regime, so the small-size cells should be reasonably independent. Note that we don't include the effect of beat-coupling discussed in Rimes & Hamilton (2006) and in Rimes et al. (2006).

For each cell in each photo- $z$  slice, we construct the two dimensional density field on  $128^2$  pixels by determining the dimensionless density contrast  $\delta$  as residual number density of the galaxies,  $\delta \equiv (n - \bar{n})/\bar{n}$ , where  $n$  is the number density of the galaxies belonging to a given pixel, and  $\bar{n}$  is the mean number density of galaxies averaged over a given cell in a given photo- $z$  slice. For the calculation of the mean number density  $\bar{n}$ , we exclude the gap regions, counting only unmasked pixels.

Then, we perform the Fourier-transformation of the two dimensional density field with the help of the FFT routine provided in the numerical recipes (Press et al. 1992) and calculate the galaxy angular power spectrum,  $C_G(l)$ , as a function of a multipole  $l$  in the Fourier space:

$$C_G(l) = \frac{N_{\text{total}}}{N_{\text{unmask}}} \langle |\tilde{\delta}(\mathbf{l})|^2 \rangle, \quad (1)$$

where  $\mathbf{l}$  is a two dimensional multipole vector representing the Fourier transform of the two dimensional angular position vector,  $N_{\text{total}}$  is the total number of pixels in a given cell (i.e.,  $N_{\text{total}} = 128^2$ ), and  $N_{\text{unmask}}$  represents the number of pixels which do not belong to the masked gap regions in a given cell. We weight the an-

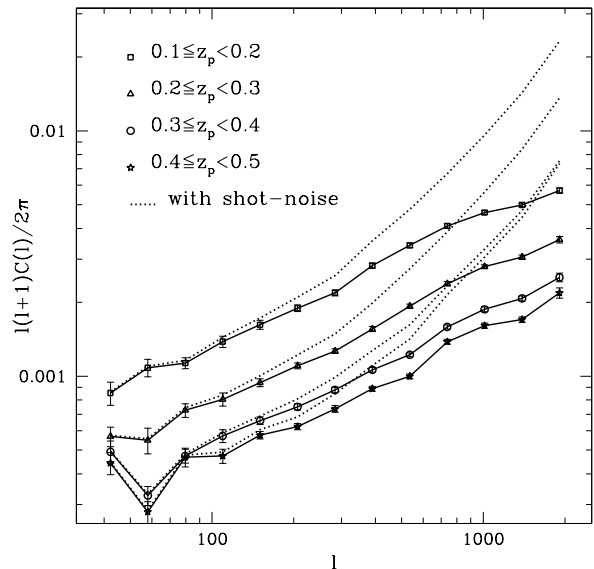


FIG. 1.— Mean angular power spectrum in a dimensionless form,  $C(l)l(l+1)/2\pi$ , averaged over 55 cells in four photo- $z$  slices:  $0.1 \leq z < 0.2$  (open squares);  $0.2 \leq z < 0.3$  (open triangles);  $0.3 \leq z < 0.4$  (open circles);  $0.4 \leq z < 0.5$  (open stars). The errors are calculated as one standard deviation in the measurement of the mean angular power spectrum. The dotted lines represent the cases before the shot-noise subtraction.

gular power spectrum by a factor of  $(N_{\text{unmask}}/N_{\text{total}})^{-1}$  to compensate for the zero powers assigned to the gap regions (Hamilton 2005). The average fraction of  $N_{\text{unmask}}/N_{\text{total}}$  is found to be  $0.92 \pm 0.02$ .

The galaxy angular power spectrum,  $C_G(l)$ , obtained by equation (1) includes the shot-noise. To eliminate the shot-noise power from a given cell to which a total of  $N_g$  galaxies belong, we construct a random sample of equal number of galaxies by generating the values of  $(\delta_I, \alpha)$  randomly. If a randomly generated galaxy is found to fall in gap regions, we regenerate  $(\delta_I, \alpha)$  till it belongs to a unmasked pixel. Using this random sample of  $N_g$  galaxies, we repeat the whole process: constructing the density field on  $128^2$  pixels, Fourier-transforming the density field, calculating the shot-noise power spectrum,  $C_{\text{SN}}(l)$ , which is also weighted by the same factor of  $N_{\text{total}}/N_{\text{unmask}}$ . Finally, we subtract the shot-noise power spectrum,  $C_{\text{SN}}$ , from  $C_G(l)$ , to obtain the shot-noise-free angular power spectrum,  $C(l)$  for all 55 cells in each photo- $z$  slice.

Figure 1 plots the mean angular power spectra,  $l(l+1)C(l)/2\pi$ , averaged over 55 cells after the shot-noise subtraction as solid lines with errors for four different photo- $z$  slices. The errors,  $\sigma_C$ , are calculated as one standard deviation in the measurement of the mean angular power spectrum averaged over 55 cells belonging to each photo- $z$  slice:  $\sigma_C \equiv \sqrt{\langle \Delta C^2(l) \rangle / (N_C - 1)}$ . For the comparison, the mean angular power spectrum,  $\bar{C}_G(l)$ , before the shot-noise subtraction is also plotted as dotted lines. As can be seen, before the shot-noise subtraction, the slope of the mean angular power spectrum changes from  $\sim 1$  to  $\sim 2$  as the multipole  $l$  increases. After the shot-noise subtraction, its slope stays constant at  $\sim 1$ .

The fluctuations of the angular power spectrum at the lowest multipole ( $l \leq 100$ ) represent the presence of the

sample variance caused by the finite size of the sky cells. Note that the amplitude of the angular power spectrum decreases with photo- $z$ , indicating the growth of powers. Here, we focus mainly on the photo- $z$  ranges of  $[0.1, 0.5)$ : At higher photo- $z$  ( $z_p \geq 0.5$ ), the shot-noise will dominate; At lower photo- $z$  ( $z_p < 0.1$ ), due to the increase of the fractional distance errors, the three dimensional analysis will be necessary to measure the power spectrum (Dodelson et al. 2001; Tegmark et al. 2002).

### 3. ANALYSIS

The angular power spectrum estimated in §2 is supposed to contain information about the amplitude of the linear power spectrum:  $C(l) = C(l; \ln A)$  where  $\ln A$  is the log of the amplitude of the linear density power spectrum. According to RH05, the Fisher information contained in the angular power spectrum about the amplitude of the linear power spectrum can be written as (Tegmark et al. 1997)

$$I \equiv - \left\langle \sum_{i,j} \frac{\partial \ln C(l_i)}{\partial \ln A} \text{Cov}_{ij}^{-1} \frac{\partial \ln C(l_j)}{\partial \ln A} \right\rangle. \quad (2)$$

Here  $(\text{Cov}_{ij}^{-1})$  represents the inverse of the covariance matrix whose components are given as

$$\text{Cov}_{ij} = \frac{\langle \Delta C(l_i) \Delta C(l_j) \rangle}{\bar{C}(l_i) \bar{C}(l_j)}, \quad (3)$$

where  $\Delta C(l) \equiv C(l) - \bar{C}(l)$  is the scatter of the angular power spectrum from the mean value at a given multipole  $l$ . As explained in detail by Rimes & Hamilton (2006), for the case of linear power spectrum,  $C^0(l)$ , from the Gaussian density fluctuations, the inverse of the covariance matrix is diagonal and equation (3) will be simplified into  $I(l_i) = n(l_i)/2$  since  $\partial \ln C(l)/\partial \ln A = 1$  and  $\text{Cov}_{ii}^{-1} = n(l_i)/2$  where  $n(l_i)$  is the number of Gaussian modes around the multipole  $l_i$ .

RH05 asserted that the amount of information preserved in the nonlinear power spectrum depends on the existence of an invertible one-to-one mapping between the linear and nonlinear regime. If such a mapping exists, then the same amount of information that the linear power spectrum contains about the initial amplitude can be extracted also from the nonlinear power spectrum. If not, then it would be difficult to determine the initial amplitudes with high precision from the nonlinear power spectrum. However, note that what we have measured in §2 is the *nonlinear galaxy power spectrum*. Unlike the nonlinear matter power spectrum that is reasonably well understood (Hamilton et al. 1991; Peacock & Dodds 1996), the non-linear clustering of galaxies is rather heuristic and usually constructed to fit the data. Although the non-linear galaxy power spectrum  $C(l)$  is known to be neither the nonlinear matter power spectrum nor the linear matter power spectrum, we set  $\bar{C}(l)$  to the linear matter power spectrum  $A C^0(l)$  for simplicity.

Using the mean angular power spectra measured at four photo- $z$  slices, we first calculate the normalized covariance matrix defined as

$$\hat{\text{Cov}}_{ij} = \frac{\langle \Delta C(l_i) \Delta C(l_j) \rangle}{\sqrt{\langle \Delta C^2(l_i) \rangle \langle \Delta C^2(l_j) \rangle}}. \quad (4)$$

Figure 3 plots the magnitudes of  $(\hat{\text{Cov}}_{ij})$  as grey scales for the cases of four photo- $z$  slices:  $[0.1, 0.2)$ ,  $[0.2, 0.3)$ ,  $[0.3, 0.4)$  and  $[0.4, 0.5)$  in the panel (a), (b), (c) and (d), respectively. As can be seen, the magnitudes of the off-diagonal elements of  $(\hat{\text{Cov}}_{ij})$  deviate from zero, which suggests the reduced amount of information than the case of Gaussian fluctuations.

Finding the inverse of the covariance matrix,  $\text{Cov}^{-1}$ , we calculate the cumulative information  $I(< l)$  contained in the SDSS angular power spectrum:

$$I(< l_k) = \sum_{j=1}^k \sum_{i=1}^k y_i \text{Cov}_{ij}^{-1} y_j, \quad k = 1, \dots, n_l \quad (5)$$

where  $\mathbf{y} \equiv (y_i)$  is a column vector whose component is all unity and  $n_l$  is the total number of multipole bins. Figure 2 plots  $I(< l)$  for the cases of four photo- $z$  slices. The errors are calculated from the 1000 times bootstrap resampling, and the dashed line represents the case of Gaussian-fluctuations. We offset slightly the horizontal positions of the four symbols at each multipole bin to show the errors clearly.

If there were a one-to-one map from the nonlinear to the linear regimes, then  $I(< l)$  obtained from SDSS catalogs would follow the dashed line. As can be seen, at low multipoles ( $l < 300$ ) it matches the Gaussian case pretty well. At higher multipole ( $l \geq 300$ ), however, we detect a clear signal of saturation in  $I(< l)$ . Note that at  $l = 2000$  the observed amount of the cumulative information is two orders of magnitude lower than expected for the case of Gaussian-fluctuations in all four photo- $z$  slices. The amplitude of the saturated region of  $I(< l)$  at high  $l$  tends to increase as photo- $z$  increases in the range of  $0.1 \leq z_p < 0.4$ , which is consistent with the numerical results of RH05. Although the amplitude becomes lower at the highest photo- $z$  slice ( $0.4 \leq z_p < 0.5$ ), the differences are within the errors. Unlike RH05, however, we could not find a sharp rise of  $I(< l)$  at nonlinear scales even when we increase the number of pixels. It might be because in the nonlinear scale the light-to-matter bias becomes important and many other complicated baryonic processes might have destroyed the information content further.

### 4. DISCUSSION AND CONCLUSION

It is worth mentioning here that what you have calculated using the SDSS data is not rigorously the information about the amplitude of the power spectrum. But, it only approximates the information in the linear power spectrum to the extent that  $d \ln P / d \ln A$  is near unity.

Our result provides an observational evidence for the previous theoretical clues that at translinear scale there is very little independent information in the matter power spectrum, consistent with light tracing matter till this scale (Rimes & Hamilton 2005, 2006; Rimes et al. 2006; Neyrinck et al. 2006; Neyrinck & Szapudi 2007). Since the multipole scales where the information saturation is detected correspond to the weak lensing regime, this result has a direct impact on the weak lensing analyses. The loss of information in the angular power spectrum would lead to increasing non-Gaussianity in sampling errors. Unlike the usual assumption adopted in most weak lensing analyses that the non-Gaussianity contribution to the sampling errors for the angular power spectrum

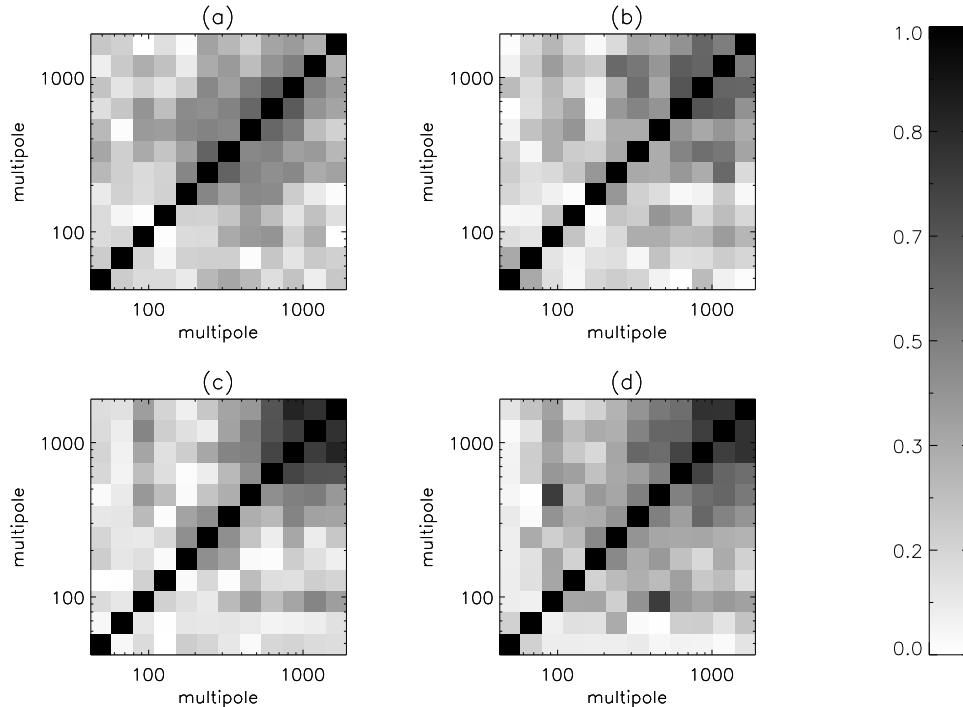


FIG. 2.— Covariance matrices of the galaxy angular power spectrum,  $C(l)$ , at four photo- $z$  slices: (a)  $0.1 \leq z_p < 0.2$ , (b)  $0.2 \leq z_p < 0.3$ , (c)  $0.3 \leq z_p < 0.4$  and (d)  $0.4 \leq z_p < 0.5$ . As done in Rimes & Hamilton (2005), the grey scale is used to represent the magnitude of the covariances.

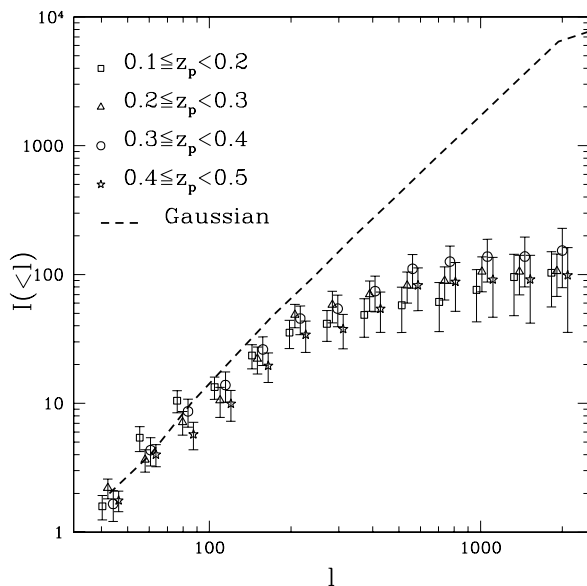


FIG. 3.— Cumulative information in the angular power spectrum for the four photo- $z$  ( $z_p$ ) slices. The dashed line represents the cumulative information in the linear angular power spectrum and the errors are calculated from 1000 times bootstrap-resampling.

is marginal (White & Hu 2000; Cooray & Hu 2001, and references therein), our result suggests that it should be quite substantial and thus the Gaussian-noise description for the lensing power spectrum should not be valid.

We thank our referee, A. Hamilton for useful suggestions. We also thank K. Kratter and S. Bonoli for their

many helps with the IDL routines. J.L. acknowledges the financial support from the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean Government (MOST, NO. R01-2007-000-10246-0).

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

## REFERENCES

- Adelman-McCarthy, J. K., et al. 2007, *ApJ*, 172, 634  
Cooray, A. & Hu, W. 2001, *ApJ*, 554, 56  
Dodelson, S., et al. 2001, *ApJ*, 572, 140  
Hamilton, A. J. S., Kumar, P., Lu, E. & Matthews, A. 1991, *ApJ*, 374, L1  
Hamilton, A. J. S. 2005, *Data Analysis in Cosmology proceedings of an International Summer School*, (Valencia: Springer-Verlag)  
Hamilton, A. J. S., Rimes, C. D., & Scoccimarro, R. 2006, *MNRAS*, 371, 1188  
Hoekstra, H. & Jain, B. 2008, *Ann. Rev. Nuclear and Particle Science*, 58, 1056  
Massey, R., et al. 2007, *Nature*, 445, 286  
Neyrinck, M. C., Szapudi, I. & Rimes, C. D. 2006, *MNRAS*, 370, L66  
Neyrinck, M. C. & Szapudi, I. 2007, *MNRAS*, 375, L51  
Rimes, C. D., & Hamilton, A. J. S. 2005, *MNRAS*, 360, L82  
Rimes, C. D., & Hamilton, A. J. S. 2006, *MNRAS*, 371, 1205  
Rimes, C. D., Hamilton, A. J. S., & Scoccimarro, R. 2006, *MNRAS*, 371, 1188  
Peacock, J. A. & Dodds, S. J. 1996, *MNRAS*, 280, L19  
Press, W. H., et al. 1992, *Numerical Recipes in Fortran* (Cambridge: Cambridge Univ. Press)  
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, *ApJ*, 480, 22  
Tegmark, M., et al. 2002, *ApJ*, 571, 191  
White, M. & Hu W. 2000, *ApJ*, 537, 1